Lecture 4: Discrete Random Variables and Distributions

MSU-STT-351-Sum-19B

Contents

In this lecture, we discuss the following:

1. Random Variables
2. Discrete Random Variables
3. Discrete Probability Distributions
4. Expected Value and Variance
5. Homework

Random Variables

Definition 1

Random Variable: A random variable X is a real-valued function on the sample space S. That is, $X : S \to R$, where R is the set of all real numbers.

Note that the value of a random variable depends on a random event.

Types of a Random Variable:

(i) A rv X is discrete if we can list its all possible values; that is, it assumes only distinct (finite or countable) values.

(ii) A rv X is called continuous if it assumes any value in a finite or infinite interval.

Example 1 Suppose an **unbiased** (that is, $P(H) = P(T) = 1/2$) coin is tossed (independently) twice and the number of heads is observed. Then the sample space is

$$S = \{HH, HT, TH, TT\}.$$

Let the rv X denote the number of heads observed. Then $X : S \to R$ and $X \in \{0, 1, 2\}$. Note X is discrete. Since the coin is unbiased, its distribution can be obtained, using independence, as

$$P(X = 0) = P(TT) = P(T)P(T) = \frac{1}{4};$$
$$P(X = 1) = P(HT) + P(TH) = \frac{1}{2};$$
$$P(X = 2) = P(HH) = \frac{1}{4}.$$

Example 2 Suppose a coin is tossed (independently) till the first head (H) occurs. Then

$$S = \{H, TH, TTH, ...\},$$

which is a countable set. Let the rv X denote the number of trials for the first H. Then $X \in \{1, 2, 3, ...\} = \{n \mid n \geq 1 \text{ and an integer}\}$. Note X is again a discrete rv.

Suppose $P\{H\} = p,\ 0 < p < 1$, and $P\{T\} = q$ so that $q = (1 - p)$. Then

$$\begin{aligned} P(X = 1) &= P\{H\} = p, \\ P(X = 2) &= P\{TH\} = P\{T\}.P\{H\} = qp, \text{ and} \\ P(X = n) &= q^{n-1}p, \quad n = 1, 2, ... \end{aligned}$$

Example 3 (Ex.5) If the sample space S is an infinite set, does it imply any r.v. on S will also assume infinitely many values? If no, give a counter example.

Solution: No. Consider the experiment in which a coin is tossed repeatedly until a H results.

Let $Y = 1$ if the experiment terminates with at most 4 tosses and $Y = 0$ otherwise. Though the sample space is infinite, the r.v. $Y \in \{0, 1\}$.

Definition: Let X be a discrete rv with $X \in \{x_n | n \geq 1\}$. Then

(i) The probability distribution or probability mass function (*pmf*) of X is $p(x_i) = P(X = x_i),\ x_i \in S$. Note $0 \leq p(x_i) \leq 1$, and $\sum_{x_i} p(x_i) = 1$.

It is the collection of all possible values of X and the corresponding probabilities. That is,

$$\begin{array}{ll} \text{Values} & : x_1, \quad x_2, \ldots \\ \text{Probabilities} & : p(x_1), p(x_2), \ldots \end{array}$$

(ii) The **cumulative distribution function** (*cdf*) of X corresponfing to *pmf* $p(x)$ is

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i), \text{ for } x \in \mathbb{R}.$$

Discrete Distributions

(iii) One use of the *cdf* is in calculating probabilities. For reals a and b with $a \leq b$,

$$P(a \leq X \leq b) = F(b) - F(a-);$$
$$P(a < X \leq b) = F(b) - F(a)$$

The notation $F(a-)$ indicates evaluation of F at the point immediately to the left of a (that is, the left limit of F at a). For most discrete random variables, F changes only at integer values n so that $F(n-) = F(n-1)$. In this case,

$$p(n) = P(X = n) = F(n) - F(n-1), \; n \in \mathbb{Z}_+ = \{1, 2, \ldots\}.$$

Example 4: Consider the Example 2, where X denotes the number of trials required for the first head H. Then its *pmf* is

$$P(X = n) = \begin{cases} q^{n-1}p & n = 1, 2, ... \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

and is called the geometric distribution, denoted by $G(p)$.

Here, $p = P(H)$, the probability of getting a head and $q = (1 - p)$.

As p changes, the probability value or the *pmf* changes, and p is called the **parameter** of the distribution.

Example 5 (Ex.11): An automobile service facility specializing in engine time-ups knows that 45% of all tune-ups are done on four-cylinder automobiles, 40% on six-cylinder automobiles, and 15% on eight-cylinder automobiles. Let X = the number of cylinders on the next car to be tuned.

(a) What is the *pmf* of X?

(b) Draw both a line graph and probability histogram for the *pmf* of part (a).

(c) What is the probability that the next car to be tuned has at least six cylinders? More than six cylinders?

Solution: (a)

x:	4	6	8
p(x):	0.45	0.40	0.15

(b)

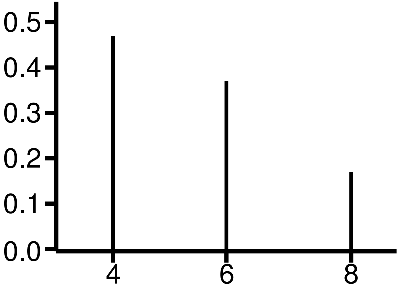

Figure: Probability Mass Function

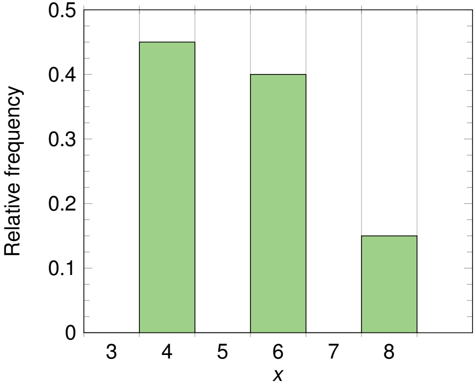

Figure: Probability Histogram

(c) $P(X \geq 6) = P(X = 6) + P(X = 8) = 0.40 + 0.15 = 0.55.$

Example 6 (Ex 15): Suppose a computer manufacturer receives computer boards **in lots of five**. Two boards are selected from each lot for inspection. Represent possible outcomes of the selection process by pairs. For example, the pair $(1, 2)$ represents the selection of boards 1 and 2 for inspection.
(a) List the ten different possible outcomes.

(b) Suppose that boards 1 and 2 are the only defective boards in a lot of five. Two boards are to be chosen at random. Define X to be the number of defective boards observed among those inspected. Find the probability distribution of X.

(c) Find the *cdf* $F(x)$ of X. (Hint: First determine $F(0) = P(X \leq 0), F(1)$ and $F(2)$; then obtain $F(x)$ for all other x.)

Solution:

(a) Note a lot $L = \{1,2,3,4,5\}$ contain five boards. Two boards are drawn at random. The sample space is

$$S = \left\{\{1,2\},\{1,3\},\{1,4\},\{1,5\},\{2,3\},\{2,4\},\{2,5\},\{3,4\},\{3,5\},\{4,5\}\right\}.$$

(b) Note $X \in \{0,1,2\}$, as it denotes the number of defective boards. Then

$$P(X=0) = p(0) = P[\{(3,4)(3,5)(4,5)\}] = \frac{3}{10} = 0.3;$$

$$P(X=2) = p(2) = P[\{(1,2)\}] = \frac{1}{10} = 0.1;$$

$$P(X=1) = p(1) = 1 - [p(0)+p(2)] = .60$$

and $P(X) = 0$ if $x \neq 0,1,2$.

(c) Note $F(x) = P(X \leq x)$. So,

$$\begin{aligned} F(0) =& P(X \leq 0) = P(X = 0) = .30 \\ F(1) =& P(X \leq 1) = P(X = 0) + P(X = 1) = .90 \\ F(2) =& P(X \leq 2) = 1. \end{aligned}$$

Hence, the *cdf* of X is

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.30, & 0 \leq x < 1 \\ 0.90, & 1 \leq x < 2 \\ 1, & 2 \leq x. \end{cases}$$

Example 7 (Ex. 24) An insurance company offers its policyholders a number of different premium payment options. For a randomly selected policyholder, let X denote the number of months between successive payments. The *cdf* of X is as follows:

$$F(x) = \begin{cases} 0, & x < 1 \\ 0.30, & 1 \leq x < 3 \\ 0.40, & 3 \leq x < 4 \\ 0.45, & 4 \leq x < 6 \\ 0.60, & 6 \leq x < 12 \\ 1, & 12 \leq x. \end{cases}$$

(a) What is the *pmf* of X?
(b) Using just the *cdf*, compute $P(3 \leq X \leq 6)$ and $P(4 \leq X)$.

Solution:

(a) Possible values of X are those values x at which $F(x)$ jumps, and $P(X = x)$ is the size of the jump at x. Thus the *pmf* of X is

x :	1	3	4	6	12
$p(x)$:	0.30	0.10	0.05	0.15	0.40

(b) Using *cdf*,

$P(3 \leq X \leq 6) = F(6) - F(3^-) = F(6) - F(1) = 0.60 - 0.30 = 0.30;$

$P(4 \leq X) = 1 - P(X < 4) = 1 - F(4-) = 1 - \{F(1) + F(3)\} = 1 - 0.40 = 0.60.$

Example 8 (Ex.18) Two fair six-sided dice are tossed independently. Let M = the maximum of the two tosses (so $M(1,5) = 5, M(3,3) = 3$, etc.).
(a) What is the *pmf* of M?
(b) Determine the *cdf* of M and graph it.

Solution: (a): Note $M \in \{1, 2, \ldots, 6\} = S$. Hence,

$$\begin{aligned} p(1) &= P(M=1) = P\Big((1,1)\Big) = \frac{1}{36}; \\ p(2) &= P(M=2) = P\Big((1,2) \text{ or } (2,1) \text{ or } (2,2)\Big) = \frac{3}{36}; \\ p(3) &= P(M=3) = P\Big(1,3) \text{ or } (2,3) \text{ or } (3,1) \text{ or } (3,2) \text{ or } (3,3)\Big) = \frac{5}{36}. \end{aligned}$$

Similarly, $p(4) = \frac{7}{36}$, $p(5) = \frac{9}{36}$ and $p(6) = \frac{11}{36}$.

(b) The *cdf* of X is

$$F(m) = \begin{cases} 0, & m < 1 \\ \frac{1}{36}, & 1 \leq m < 2 \\ \frac{4}{36}, & 2 \leq m < 3 \\ \frac{9}{36}, & 3 \leq m < 4 \\ \frac{16}{36}, & 4 \leq m < 5 \\ \frac{25}{36}, & 5 \leq m < 6 \\ 1, & 6 \leq m. \end{cases}$$

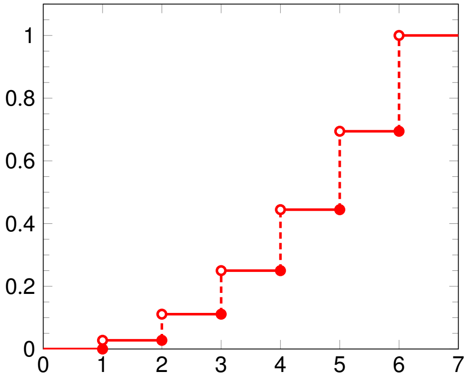

Figure: The cdf of M

Expected Value and Variance

Definition 2

Let X be a discrete random variable with *pmf* $p(x)$. The expected value or the mean value of X, denoted by $\mu_x = E(X)$, is

$$\begin{aligned}\mu_x &= x_1 P(X = x_1) + x_2 P(X = x_2) + \dots \\ &= \sum_{i=1}^{\infty} x_i p(x_i).\end{aligned}$$

The variance of X, denoted by $Var(X) = \sigma_x^2$ is

$$\begin{aligned}\sigma_x^2 &= E(X - \mu_x)^2 \\ &= (x_1 - \mu_x)^2 P(X = x_1) + (x_2 - \mu_x)^2 P(X = x_2) + \dots \\ &= \sum_{i=1}^{\infty} (x_i - \mu_x)^2 p(x_i).\end{aligned}$$

Expected Value and Variance

Sometimes, it is simply written as

$$\mu_x = E(X) = \sum_i x_i P(X = x_i);$$
$$\sigma_x^2 = Var(X) = \sum_i (x_i - \mu_x)^2 P(X = x_i).$$

The standard deviation (SD) of X is $\sigma_x = \sqrt{\sigma_x^2}$. (the positive square root).

Indeed, the expected value of $h(X)$, a function of X, is defined as

$$E(h(X)) = \sum_i h(x_i)p(x_i).$$

A shortcut formula for σ_x^2

The variance of X can be computed easily using the formula

$$\sigma_x^2 = E(X^2) - [E(X)]^2 = \sum_i x_i^2 p(x_i) - \mu_x^2.$$

Rules for Mean and the Variance
The following relations can be easily established:

$$\begin{aligned}
a)\ \ \mu_{aX+b} &= E(aX+b) = a\mu_X + b;\\
b)\ \ \mu_{X+Y} &= E(X+Y) = \mu_X + \mu_Y;\\
c)\ \ \sigma^2_{aX+b} &= Var(aX+b) = a^2\sigma_X^2;\\
d)\ \ \sigma_{aX+b} &= |a|\sigma_X
\end{aligned}$$

for **any** random variables X and Y.

Definition 3 (Independence of RVs)

The random variables X and Y are independent if

$$P(X \in A; Y \in B) = P(X \in A)P(Y \in B),$$

for all events A and B.

If X and Y are independent rvs, then

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y; \text{ but}$$

$$\sigma_{X+Y} = \sqrt{(\sigma^2_X + \sigma^2_Y)} \neq \sigma_X + \sigma_Y$$

Example 9 (Ex 29) The *pmf* for X = the number of major defects on a randomly selected appliance of a certain type is

x:	0	1	2	3	4
p(x):	0.08	0.15	0.45	0.27	0.05

Compute the following: (a) $E(X)$
(b) $V(X)$ directly from the definition
(c) The standard deviation of X
(d) $V(X)$ using the shortcut formula

Expected Value and Variance

Solution:

(a):

$$E(X) = \sum_{i=0}^{4} x_i p(x_i)$$
$$=(0)(.08) + (1)(.15) + (2)(.45) + (3)(.27) + (4)(.05) = 2.06.$$

(b):

$$V(X) = \sum_{i=0}^{4} (x_i - 2.06)^2 p(x_i)$$
$$=(0 - 2.06)^2(.08) + \ldots + (4 - 2.06)^2(.05) = .9364.$$

(c): $\sigma_x = \sqrt{(0.9364)} = 0.9677.$

(d): Using the shortcut formula,

$$V(X) = \sum_{i=0}^{4} x_i^2 p(x_i) - (2.06)^2 = 5.1800 - 4.2436 = 0.9364.$$

Example 10 (Ex 37) Suppose n candidates for a job have been ranked $1, 2, 3, \ldots, n$. Let X = the rank of a randomly selected candidates, so that X has *pmf*

$$P(X = x) = p(x) = \begin{cases} \frac{1}{n}, & x = 1, 2, ..., n \\ 0, & \text{otherwise.} \end{cases}$$

This is called the **discrete uniform distribution**.

Compute $E(X)$ and $V(X)$, using the shortcut formula. [Hint: The sum of the first n positive integers is $n(n+1)/2$, whereas the sum of their squares is $n(n+1)(2n+1)/6$.]

Expected Value and Variance

Solution:

$$E(X) = \sum_{k=1}^{n} k\Big(\frac{1}{n}\Big) = \Big(\frac{1}{n}\Big)\sum_{k=1}^{n} k = \frac{1}{n}\Big[\frac{n(n+1)}{2}\Big] = \frac{n+1}{2},$$

$$E(X^2) = \sum_{k=1}^{n} k^2\Big(\frac{1}{n}\Big) = \frac{1}{n}\Big[\frac{n(n+1)(2n+1)}{6}\Big] = \frac{(n+1)(2n+1)}{6}.$$

So,

$$Var(X) = \frac{(n+1)(2n+1)}{6} - \Big(\frac{n+1}{2}\Big)^2 = \frac{n^2-1}{12}.$$

Example 11: Find the expected value and SD of random variable X with

x:	0	1	2
P(X=x):	0.2	0.4	0.4

Solution:
(The mean and the variance are given by:

$$\begin{aligned} E(X) &= \mu_X = 0(0.2) + 1(0.4) + 2(0.4) = 1.2 \\ V(X) &= (0 - 1.2)^2(0.2) + (1 - 1.2)^2(0.4) + (2 - 1.2)^2(0.4) \\ &= 0.56 \end{aligned}$$

Hence,

$$\sigma_X = \sqrt{0.56} = 0.748.$$

Exercise 38: Let X denote the outcome when a fair die rolled once. If before the die is rolled you are offered either $(1/3.5)$ dollars or $h(X) = 1/X$ dollars, after the die is rolled, would you accept the guaranteed amount or would you gamble? [Note: It is **not** generally true that $1/E(X) = E(1/X)$]

Solution: Note that if you gamble, your average gain is

$$E[h(X)] = E(\frac{1}{X}) = \sum_{i=1}^{6}(\frac{1}{x_i})p(x_i) = \frac{1}{6}\sum_{i=1}^{6}\frac{1}{x_i} = 0.408.$$

If you do not gamble, you get $\frac{1}{3.5} = 0.286$. Hence, you expect to win more if you gamble.

Example 12: Kids: A couple plans to have children until they get a girl, but they agree they will not have more than three children even if all are boys.

a) Create a probability model for the number of children they'll have.

b) Find the expected number of children.

c) Find the expected number of boys they'll have.

d) Find the SD of the number of children the couple may have.

Solution:

Outcomes	Prob.
G	0.5
BG	(0.5)(0.5)=0.25
BBG	(0.5)(0.5)(0.5)=0.125
BBB	(0.5)(0.5)(0.5)=0.125

Let X=Number of children. Then the *pmf* of X is
(a)

x:	1	2	3
p(x):	0.5	0.25	0.25

(b)

$$E(X) = 1(.5) + 2(.25) + 3(.25) = 1.75$$

(c) Let Y=Number of boys. Then the *pmf* of Y is

y:	0	1	2	3
p(y):	0.5	0.25	0.125	0.125

$$\begin{aligned}(d) : \sigma_X^2 &= (1 - 1.75)^2(.5) + (2 - 1.75)^2(.25) + (3 - 1.75)^2(.25) \\ &= 0.6875.\end{aligned}$$

Hence, $\sigma_X = \sqrt{0.68755} = 0.829$.

Example 13: Given **independent** random variables with means and standard deviations as shown, find the mean and standard deviation of each of these variables:

(a) $X - 20$
(b) $0.5Y$
(c) $X + Y$
(d) $X - Y$

	Mean	SD
X:	80	12
Y:	12	3

Expected Value and Variance

Solution:

(a)
$$\mu_{X-20} = \mu_x - 20 = 80 - 20 = 60$$
$$\sigma_{X-20} = \sigma_X = 12.$$

(b)
$$\mu_{(0.5)Y} = (0.5)\mu_Y = (0.5)(12) = 6$$
$$\sigma_{(0.5)Y} = |0.5|\sigma_Y = (0.5)(3) = 1.5$$

(c)
$$\mu_{X+Y} = \mu_X + \mu_Y = 80 + 12 = 92$$
$$\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2} = \sqrt{12^2 + 3^2} = \sqrt{153} = 12.369$$

d)
$$\mu_{X-Y} = \mu_X - \mu_Y = 80 - 12 = 68$$
$$\begin{aligned}\sigma_{X-Y} = &\sqrt{\sigma_X^2 + \sigma_{-Y}^2} \\ &= \sqrt{\sigma_X^2 + (-1)^2\sigma_Y^2} \\ &= \sqrt{\sigma_X^2 + \sigma_Y^2} \\ &= 12.369.\end{aligned}$$

Homework

Homework:

Section 3.1: 1, 7, 9

Section 3.2: 11, 13, 18, 23

Section 3.3: 33, 35, 39, 43

−− END OF LECTURE 4 −−